Tensions in data sharing and reuse
The perspective of researchers

Kathleen Gregory | @gregory_km
University of Vienna
Scholarly Communications Lab, University of Ottawa
Overlapping practices

Managing data,
Sharing data,
Discovering data,
Reusing data,
Citing data…
4 Tensions
In practice, policy, infrastructure

Accounting for data diversity
New and traditional practices
Balancing practice and policy
Data management is work.
Accounting for data diversity

Tension One
Diversity of (meta)data

One person’s signal is another person’s noise

Data do not stand alone.

Beyond disciplines

Diversity within & across disciplines

Belong to multiple ‘data communities’ with own norms

New and traditional practices

Tension Two
Citing data

Variation in how reference data
Tied to long-standing community practices

How do you cite or reference data?

- Mentioning data in captions, figures or tables

- Mentioning data in acknowledgements, appendix, etc
  - Never: 28.6, Rarely: 23.2, Sometime: 23.6, Often: 13.1, Always: 11.6

- Mentioning data in body of text

- Mentioning data in footnote
  - Never: 43.6, Rarely: 21.8, Sometime: 19.5, Often: 8, Always: 7.1

- Including a citation to a related paper in reference lists
  - Never: 5.6, Rarely: 14.3, Sometime: 23.8, Often: 51.1

- Including a citation to data in reference lists
  - Never: 9.3, Rarely: 6.3, Sometime: 10.6, Often: 19.9, Always: 51.4

Weighted percentages of respondents to this question (n=2,026)

Citing data

Challenge of new standards?

Are you aware of & do you use these standards for citing data?

- Citation style guides
  - I am aware of and use this recommendation: 58.7%
  - I am aware of and do not use this recommendation: 8.6%
  - I am unaware of this recommendation: 32.7%

- Journals or publishers*
  - I am aware of and use this recommendation: 77.5%
  - I am aware of and do not use this recommendation: 5.8%
  - I am unaware of this recommendation: 16.7%

- Scientific societies
  - I am aware of and use this recommendation: 28.1%
  - I am aware of and do not use this recommendation: 12.1%
  - I am unaware of this recommendation: 59.8%

- Data repositories*
  - I am aware of and use this recommendation: 42.3%
  - I am aware of and do not use this recommendation: 14.6%
  - I am unaware of this recommendation: 43.1%

- DataCite*
  - I am aware of and use this recommendation: 19.1%
  - I am aware of and do not use this recommendation: 8.1%
  - I am unaware of this recommendation: 72.8%

Weighted percentages of respondents to this question (n=2,492)
Balancing practice and policy

Tension Three
Practice vs policy

When do we meet researchers where they are?
Data management is work.

Tension Four
This takes work.
Detailed, iterative work
Duplicative work
If we add to researchers’ plates, what can get taken away?

Well, yes…but what about just more time?
To recap

Accounting for data diversity
New and traditional practices
Balancing practice and policy
Data management is work.
Thank you.

Questions?

Kathleen Gregory | @gregory_km
kathleen.gregory@univie.ac.at


